



Individualizing the WOMAC function subscale: Incorporating patients' priorities for improvement to measure functional impairment in hip or knee osteoarthritis

Raphaële Seror, Florence Tubach, Gabriel Baron, Bruno Falissard, Isabelle Logeart, Maxime Dougados and Philippe Ravaud

Ann Rheum Dis published online 17 Aug 2007;
doi:10.1136/ard.2007.074591

Updated information and services can be found at:
<http://ard.bmj.com/cgi/content/abstract/ard.2007.074591v1>

These include:

Rapid responses

You can respond to this article at:
<http://ard.bmj.com/cgi/eletter-submit/ard.2007.074591v1>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

Online First contains unedited articles in manuscript form that have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Online First articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Online First articles must include the digital object identifier (DOIs) and date of initial publication.

To order reprints of this article go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Annals of the Rheumatic Diseases* go to:
<http://journals.bmj.com/subscriptions/>

Individualizing the WOMAC function subscale:
Incorporating patients' priorities for improvement to measure
functional impairment in hip or knee osteoarthritis

Raphaelle Seror ¹, Florence Tubach ¹, Gabriel Baron ¹, Bruno Falissard ², Isabelle Logeart ³,
Maxime Dougados ⁴, Philippe Ravaud ¹

(1), AP-HP, Hôpital Bichat, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Paris, France; INSERM, U738, Paris, France; Université Paris 7 Denis Diderot, Paris, France

(2) AP-HP, Hôpital Paul Brousse, Département de santé publique, Villejuif, France ; INSERM, U669, Paris, France; Université Paris XI, France.

(3) Merck Sharp & Dohme, Chibret Laboratories, Paris, France.

(4) AP-HP, Hôpital Cochin, Service de Rhumatologie B, Paris, France; Université Paris-Descartes, Paris, France.

Correspondence and reprint requests to:

Dr Raphaële Seror

Département d'Epidémiologie, Biostatistique et Recherche Clinique, INSERM, U738, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France.

Email : raphaele.se@gmail.com

KEY WORDS: WOMAC, osteoarthritis, individualization, patients' perspectives, patient reported outcome.

WORD COUNT: 3102

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article to be published in Annals of the Rheumatic Diseases editions and any other BMJ PGL products to exploit all subsidiary rights, as set out in our licence <http://ard.bmjournals.com/ifora/licence.pdf>"

ABSTRACT

Objective. Recommended outcome measures in osteoarthritis are standardized scales identical for each patient. As patient-specific scales are of increasing interest to consider patient's priorities in outcome assessment, this study aims to validate individualized forms of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) function subscale.

Patients and Methods. WOMAC function subscale data were prospectively obtained from 1218 outpatients with hip or knee osteoarthritis requiring nonsteroidal anti-inflammatory drugs. Patients also rated the importance to get rid of disability in each activity of the WOMAC function subscale and select the 5 activities they considered the most important to be improved upon. After treatment, patients again completed the WOMAC function subscale. Several individualization methods were evaluated: methods whereby the score of each item is multiplied by, or added to, its importance and methods based on the 5 most important activities (WOMAC top 5). Psychometric properties of individualized scales were compared to those of the WOMAC function subscale.

Results. Missing data rate was 11%, 13% and 2% for the WOMAC function, its individualized forms and the WOMAC top 5, respectively. Combining severity and importance of each item did not improve scales' properties. The WOMAC top 5 was the most responsive scale (standardized response mean: 0.96 vs 0.80, $p < 0.0001$).

Conclusion. Because of its better responsiveness, ease of use, low missing data rate and ability to highlight patients' priorities, the WOMAC top 5 could be an interesting tool in therapeutic evaluation in hip or knee osteoarthritis.

INTRODUCTION

Patient-reported outcomes are increasing in interest in clinical practice and clinical research. In the setting of osteoarthritis (OA), a core set of outcome measures to be considered for phase III trials has been defined by the OMERACT (Outcome Measure in Rheumatology Clinical Trial); 3 domains should systematically be included: pain, physical function and patient global assessment.[1] The function subscale of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) is a valid, reliable, and responsive measure of functional impairment in hip and knee OA.[2-4] It is the most widely used condition-specific index, and the subscale is recommended to be included in all hip and knee OA trials.[1] However, the WOMAC function subscale is a standardized instrument that involves a fixed number of identical items, all having the same weight on the final score. Although the mean score is a measure of global functional impairment at the group level, the measure does not consider the variability in the importance to patients to be able to perform a particular activity. For example, the ability to climb stairs should be of poor importance in a patient who always takes a lift, or getting out of the bath should be of no importance for a patient who always takes showers.

Questionnaires that principally focus on each patient's priorities, so-called patient-specific or individualized instruments, have been developed and/or used in rheumatology.[5-12] These scales identify relevant issues at the individual level and allow the evaluation to focus on what is important to each patient. Some have shown better sensitivity to change than classical instruments.[6, 7, 13] But a few patient-specific scales have been applied in hip or knee disorders [7, 14] and none in lower limb OA requiring a medical intervention. As patients' perspectives are matters of increasing importance,[15, 16] this prospective study aimed to develop and validate individualized scales derived from the WOMAC function subscale by several methods of individualization assessing functional impairment in patients with hip or knee OA and to compare their psychometric properties to those of the WOMAC function subscale.

PATIENTS AND METHODS

Study population

Data were obtained from a prospective cohort study of 4 weeks' duration involving 1362 outpatients with hip (n=343) or knee (n=1019) OA as defined by the American College of Rheumatology.[17, 18] Between April 12 and July 31 2002, patients were recruited by 399 French rheumatologists in private practice. Each rheumatologist was required to include 3 patients with knee OA and 1 with hip OA. To be included in the study, patients had to experience pain from the OA (≥ 30 mm on a visual analogue scale (VAS) [0-100 mm]). Each patient gave an informed consent. All patients initially visited the rheumatologist in charge of their case and inclusion could begin with the onset of a nonsteroidal anti-inflammatory drug (NSAID) or with a switch from one NSAID to another. A final visit to the same rheumatologist was scheduled 4 weeks later.

Measurement

At baseline visit, patients completed 3 self-administered questionnaires (figure 1):

- The French-Canadian version of the WOMAC function subscale (5-point Likert version),[2] also termed as “severity questionnaire” in this paper, which is a 17-item scale addressing the degree of difficulty in accomplishing 17 activities of daily life.
- The “importance questionnaire”: patients had to rate how important it was to them to get rid of disability in each activity addressed by the WOMAC function items (from not important at all, to extremely important). Patients were randomly assigned to 3 groups according to different response modalities for rating that questionnaire: 5-point (1-5) Likert scale (Likert 5 group), 3-point (1-3) Likert scale (Likert 3 group), or VAS [0-100 mm] (VAS group).
- The “preference questionnaire”: patients had to select the 5 items of the WOMAC function they considered the most important by answering to the following question: “Could you choose from the 17-item list, the 5 you consider the most important to be improved upon?”

Along these measures patients also assessed their pain and global disease activity on a VAS [0-100 mm]. And practitioners assessed each patient’s global disease activity on a VAS [0-100 mm].

At final visit, patients again completed the WOMAC function subscale.[2]

To assess the test–retest reliability, a subsample of 93 patients, all from the Likert 5 group, were asked to complete, and send by mail, the WOMAC function subscale, the “importance” and the “preference” questionnaires again within 48 hours, before initiating NSAID therapy.

Methods of individualization

Several methods for individualization were used:

(1) Individualized scales based on the importance questionnaire rated with a 5- or 3-point Likert scale or VAS.

- (a) With multiplicative methods: for each item the severity score was multiplied by the importance score,
- (b) With additive methods: for each item the severity score was added to the importance score.

These scales were derived from both the WOMAC function subscale (17 items) and the WOMAC short-form subscale (8 items).[19]

(2) The WOMAC top 5, based on the preference questionnaire, including the 5 most important items to each patient. Thus, the items of the WOMAC top 5 are not the same for all patients.

Statistical analysis

Because the psychometric performances of each scale were compared to those of the WOMAC function subscale, only patients who had completed this scale with no missing data at the baseline visit (n=1218) were involved in the development of the individualized forms (figure 1). Each scale was linearly transformed to a 0 to 100 scale, with a score of 0 indicating no disability and 100 indicating maximum possible disability. For each item, Spearman’s rho correlation coefficients between severity score and importance score were obtained.

Psychometric properties of each scale were evaluated, and properties of individualized scales were compared to those of the WOMAC function subscale.

- Construct validity was assessed by Spearman's rho correlation coefficient between the scores of each scale and that of the WOMAC function subscale. We examined divergent validity by the use of Spearman's correlation coefficients between scores of individualized scales and other measures applied in this study (pain, patient and practitioner global assessment of disease activity).
- Internal consistency was assessed when estimable (for fixed-item scales), with Cronbach's alpha coefficient.[20] Estimation of confidence intervals and comparisons of Cronbach's alpha coefficients involved use of bootstrapping methods, with 1000 replications.[21]
- Test-retest reliability was assessed with the intraclass correlation coefficient (ICC). ICC values vary from 0 (totally unreliable) to 1 (perfectly reproducible); an $ICC \geq 0.75$ is regarded as excellent.[22] ICC confidence intervals were estimated with bootstrapping methods, with 1000 replications.[21] ICCs were compared with the likelihood ratio test.[23]
- Responsiveness was assessed by the standardized response mean (SRM). SRM is the mean change in score between the baseline and the final visit divided by the standard deviation of the change in score.[24] SRM values can be considered large (>0.8), moderate (0.5-0.8) or small (<0.5).[25, 26] SRM confidence interval estimations and SRM comparisons involved use of bootstrapping methods, with 1000 replications.[21] Because final scores were calculated using the baseline importance questionnaire, SRM of scales using additive methods are arithmetically the same as the scale from which they are derived, so comparisons were not performed.

Statistical analyses involved use of the SAS release 9.1 and R release 2.2.1 statistical software packages.

RESULTS

At baseline, the missing data rates were 11% (n=144), 13% (n=174) and 2% (n=21) for the WOMAC function subscale, its individualized derived forms and the WOMAC top 5, respectively. Baseline characteristics of the 1218 patients involved in analyses are reported in table 1 and were similar to those of patients with incomplete WOMAC function subscale data (n=144). Among these 144 patients, one patient did not complete the questionnaire, 71 (49.3%), 53 (36.8%), 11 (7.6%) and 8 (5.7%) patients did not respond to 1, 2, 3 or more than 3 items. Baseline characteristics of patients in the Likert 5, Likert 3 and VAS groups did not differ.

Mean scores for each scale and mean changes in score over the 4-week period are reported in table 2. Neither the WOMAC function subscale nor the individualized scales had a substantial ceiling or floor effect. For each item, the severity and importance scores were significantly correlated for each group (rho ranging from 0.34 to 0.67). However, for each activity of the WOMAC function subscale, some patients with a low severity score rated the item as very important, whereas a few patients with a high severity score for a given activity rated the item as being of little importance, except for item 13 ("getting in/out of the bath") and item 16 ("performing heavy domestic duties") where some patients with high severity score rated the item as being of little importance (figure 2).

Construct validity

All individualized scales were highly convergent with the WOMAC function subscale ($\rho \geq 0.75$). Individualized scales involving all 17 items correlated more strongly with the WOMAC function subscale than shorter scales such as WOMAC top 5 or scales derived from the WOMAC short form. However, the additive scale using VAS for rating of importance was less correlated with the WOMAC function subscale than all other scales. Lower correlations ($\rho < 0.5$) were obtained between all scales, measuring functional status, and pain and global assessment of activity (table 3).

Internal consistency

Cronbach's alpha coefficients of individualized scales involving all 17 items did not significantly differ from that of the WOMAC function subscale and ranged from 0.91 to 0.94 (table 3). For scales involving 8 items (derived from the WOMAC short form), Cronbach's alpha coefficients were significantly lower, ranging from 0.82 to 0.86. For the WOMAC top 5, as the items involved in the scale were different for each patient, Cronbach's alpha coefficient was not estimable.

Test-retest reliability

Among the 93 patients for whom test-retest reliability was assessed, necessary data to compute the WOMAC function subscale, its individualized forms and the WOMAC top 5 were obtained for 71 (76%), 64 (69%) and 78 (83%) patients, respectively. The mean change in score between test and retest were of 4.76 ± 11.46 , 4.24 ± 9.12 , 4.24 ± 9.23 and 6.42 ± 16.48 for the WOMAC function subscale, its individualized form with multiplicative method and additive methods and the WOMAC top 5, respectively. The ICCs of individualized scales (long and short forms) did not significantly differ from that of the WOMAC function subscale (table 3) and were > 0.75 , except for the WOMAC top 5 ($\text{ICC} = 0.58$).

Responsiveness

The SRM of the WOMAC function in the overall population was 0.80 (table 3). The SRM of the WOMAC function subscale was 0.78, 0.86 and 0.77 for the Likert 5, Likert 3 and VAS groups, respectively. The SRM of the 17-item individualized forms with multiplicative methods did not significantly differ from that of the WOMAC function subscale, except for scale using VAS for rating importance (0.85 vs 0.77, $p = 0.01$). The WOMAC short form showed better responsiveness than the WOMAC function subscale (0.84 vs 0.80, $p = 0.002$). However, the best sensitivity to change was obtained with the WOMAC top 5 (0.96 vs 0.80, $p < 0.0001$).

Sensitivity analyses

Where there were less than 3 item missing, WOMAC function scores were computed with imputation of missing data, with the average value of the subscale, as recommended in the WOMAC user's guide.[27] Analyses performed in this population of 1353 patients gave similar results as those obtained with data from 1218 patients, leading to the same conclusions.

DISCUSSION

This study aimed to validate individualized measures of function impairment in hip or knee OA developed from the patient's perspective by highlighting each patient's priorities about functional improvement. All scales were derived from the WOMAC function subscale. Adding the measure of "importance" to the measure of "severity" provided complementary information. However scales combining the severity and importance of each item did not have better properties than the WOMAC function subscale, except with VAS used for rating of importance. The WOMAC top 5 was the best scale in term of responsiveness and missing data rate.

Psychometric properties of the WOMAC function subscale found in this study are in accordance with those previously described and validated.[3, 4, 28-31] This study enrolled a large sample of patients with a wide range of severity of disease, which might allow the conclusions to be applied to a wide spectrum of patients with hip or knee OA.

Because no consensus exists concerning the best way to individualize functional status instruments, we have evaluated 2 of the main possible individualization methods. In the first method, pre-specified items, identical for each patient, are preserved in the final score; individualization involves combining for each item rating of "severity" and "importance".[7] In the second method, individualization is based on a selection process, patients are asked to specify or choose a limited (or not) number of areas they considered the most in need of improvement.[6, 9, 32] Then, patients are followed in terms of only these selected items. Moreover, both methods could be combined, for example, by rating the importance of the selected items.

In this study, the WOMAC top 5 had a low rate of missing data and a good responsiveness. In addition, this scale is probably the most patient-specific scale because the selection process ensures inclusion of only clinically relevant activities to each patient. Furthermore, a SRM of 0.96 is large in a study involving pharmacological therapy intervention (NSAIDs); greater values of responsiveness had been found mainly when the intervention was a surgical joint replacement and when response was integrated over multiple observed time points.[4, 33] The WOMAC top 5 displayed fair reliability. However, the ICC exceeded 0.5, which is the minimal necessary precondition for appropriate application of a change score and evaluation of responsiveness.[34] One hypothesis to explain the smaller ICC observed for the WOMAC top 5 could be due to the within-patient change during the first 48 hours. However this point is unlikely to explain entirely the phenomenon, as shown in the analysis of the variation in score in these patients during this period that was moderate (mean change of 6.42 for the top5 versus 4.24 to 4.76 for other scales) and in the graphical analysis of Bland and Altman plots that did not reveal major systematic change in score (data not shown). Guyatt *et al.* distinguished 2 kinds of measurement instruments: discriminative instruments, which measure the difference between subjects, and evaluative instruments, which measure change over time and treatment effects.[35, 36] In the setting of OA, outcome measures such as the WOMAC function subscale or its individualized forms, are primarily assumed to be efficient for evaluative purposes. The key issue in developing evaluative instruments is to improve their responsiveness, to allow reducing the sample sizes when designing clinical trials or developing tools for more precise treatment comparison. From this assumption, the most valuable scale is the WOMAC top 5.

Shorter tools such as WOMAC top 5 and WOMAC short form could be of interest in terms of feasibility (ease of use, time for completion, lower missing data rate) but also relevance of content. Actually, some activities addressed in the WOMAC function subscale

(4/17) are not performed by 5% to 30% of patients in their daily life [37] and thus are not relevant to them. Among these 4 items, 3 have been excluded in the WOMAC short form.[19] These items generated more missing data than the others and, if numerous, generated higher missing data rate for the overall score. The WOMAC top 5 reduces these 2 drawbacks by including only relevant activities for each patient. These findings support the fact that in case report forms, shorter scales could result in better data quality.

The importance of improving ability to perform an activity was closely related to level of disability in that particular activity (figure 2). For each activity, when rating severity, patients may also take into account the importance of being able to perform it, not just their ability to perform it. These findings could explain why most of scales developed with methods combining, for each item, severity and importance showed psychometric performances not significantly different from those of the scale from which they were derived (long or short form). However, data showed that overlap between the rating of severity and importance was incomplete (figure 2). Thus, in clinical practice, assessment of importance for improvement in each area of function may provide complementary information to assessment of functional status alone. Moreover, such patient-centred approach might reinforce the patient-physician relationship in clinical practice.[38]

We evaluated the impact of the response modalities (Likert 5, Likert 3 and VAS) used for rating importance. In most studies comparing values of VAS or multiresponse-mode scales,[2, 33, 39-41] both scales had similar results. Some authors encourage the use of the Likert scale [41-43] because of its ease of use and interpretation. Others encourage the use of the VAS [26] because of its better precision and sensitivity to change.[44, 45] We found that scales involving VAS were more sensitive to change. However, the increase in responsiveness was small compared to the complexity addressed by this method in data management.

The use of individualized versions of WOMAC function subscale, we herein developed, enables highlighting patients' concerns not only in therapeutic evaluation but also in identification of patients' priorities for improvement in clinical practice. Nevertheless, in our study, the use of a pre-established list of items did not offer patients the opportunity to provide supplemental items of relevance to them, which might be a limitation to our individualization methods, but is of more practical use particularly in case reported forms in the setting of clinical trials. These scales allow for determining whether each patient's priorities, in terms of functional improvement, are attained, the priorities having been established before treatment initiation. However, patient's priorities may change over time, due to response shift or, for instance, with improvement deterioration or change in physical environment. In this study, we did not investigate these changes. Another potentially important question is to determine the optimal number of item to increase responsiveness with preservation of measure precision. This has not been investigated in this study were patients had to select exactly 5 items, but this could be an interesting topic for future research.

Individualization by combining severity and importance for each WOMAC function item did not improve scale psychometric performances but provided complementary information on patients' priorities that could be relevant in clinical practice. Among all scales, because of its better sensitivity to change, ease of use, lower rate of missing data and better reflection of patients' concerns, the WOMAC top 5 could be an interesting tool in therapeutic evaluation and decision making for patients with hip or knee OA. It could be easily included in OA trials and used in addition to the WOMAC function subscale by adding only the preference questionnaire. The WOMAC top 5 requires further validation in independent samples of subjects from the target population.

Acknowledgments

We thank all the rheumatologists who recruited patients for this study. This study was supported by a by an unrestricted grant from Merck, Sharp & Dohme Chibret Laboratories, France.

CAPTIONS OF FIGURES

Figure 1: Flow-chart

Figure 2: Patterns of “jittered” plots of severity against importance scores, for items “descending stairs”, “standing” and “getting in/out of the bath” for the VAS group.

REFERENCES

- 1 Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol.* 1997;**24**(4):799-802.
- 2 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;**15**(12):1833-40.
- 3 Bellamy N, Kean WF, Buchanan WW, Gercz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC Osteoarthritis Index. *J Rheumatol.* 1992;**19**(1):153-9.
- 4 McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum.* 2001;**45**(5):453-61.
- 5 Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. *J Clin Epidemiol.* 2005;**58**(8):791-801.
- 6 Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire--an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol.* 1987;**14**(3):446-51.
- 7 Wright JG, Young NL. The patient-specific index: asking patients what they want. *J Bone Joint Surg Am.* 1997;**79**(7):974-83.
- 8 Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. *Can J Occup Ther.* 1990;**57**(2):82-7.
- 9 Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life. The Patient-Generated Index. *Med Care.* 1994;**32**(11):1109-26.
- 10 O'Boyle CA, McGee H, Hickey A, O'Malley K, Joyce CR. Individual quality of life in patients undergoing hip replacement. *Lancet.* 1992;**339**(8801):1088-91.
- 11 Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. *Arthritis Rheum.* 1995;**38**(11):1568-80.
- 12 Duffy CM, Arsenaault L, Duffy KN, Paquin JD, Strawczynski H. The Juvenile Arthritis Quality of Life Questionnaire--development of a new responsive index for juvenile rheumatoid arthritis and juvenile spondyloarthritis. *J Rheumatol.* 1997;**24**(4):738-46.
- 13 Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol.* 1997;**50**(3):239-46.
- 14 Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, *et al.* The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction. *Phys Ther.* 1997;**77**(8):820-9.
- 15 Carr AJ, Higginson IJ. Are quality of life measures patient centred? *Bmj.* 2001;**322**(7298):1357-60.
- 16 Wright JG. Evaluating the outcome of treatment. Shouldn't We be asking patients if they are better? *J Clin Epidemiol.* 2000;**53**(6):549-53.
- 17 Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al.* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum.* 1986;**29**(8):1039-49.
- 18 Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum.* 1991;**34**(5):505-14.
- 19 Tubach F, Baron G, Falissard B, Logeart I, Dougados M, Bellamy N, *et al.* Using patients' and rheumatologists' opinions to specify a short form of the WOMAC function subscale. *Ann Rheum Dis.* 2005;**64**(1):75-9.
- 20 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;**16**:297-334.
- 21 Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

- 22 Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol.* 2003;**56**(8):730-5.
- 23 Giraudeau B, Porcher R, Mary JY. Power calculation for the likelihood ratio-test when comparing two dependent intraclass correlation coefficients. *Comput Methods Programs Biomed.* 2005;**77**(2):165-73.
- 24 Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;**28**(7):632-42.
- 25 Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;**53**(5):459-68.
- 26 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 2nd ed ed. New York: Oxford University Press, 1996.
- 27 Bellamy N. WOMAC osteoarthritis index: a user's guide, IV. London. 2000.
- 28 Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF, Kraag GR, *et al.* A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. *J Rheumatol.* 1993;**20**(6):999-1004.
- 29 Grace D, Rogers J, Skeith K, Anderson K. Topical diclofenac versus placebo: a double blind, randomized clinical trial in patients with osteoarthritis of the knee. *J Rheumatol.* 1999;**26**(12):2659-63.
- 30 Zhao SZ, McMillen JJ, Markenson JA, Dedhiya SD, Zhao WW, Osterhaus JT, *et al.* Evaluation of the functional status aspects of health-related quality of life of patients with osteoarthritis treated with celecoxib. *Pharmacotherapy.* 1999;**19**(11):1269-78.
- 31 Peloso PM, Bellamy N, Bensen W, Thomson GT, Harsanyi Z, Babul N, *et al.* Double blind randomized placebo control trial of controlled release codeine in the treatment of osteoarthritis of the hip or knee. *J Rheumatol.* 2000;**27**(3):764-71.
- 32 Bell MJ, Bombardier C, Tugwell P. Measurement of functional status, quality of life, and utility in rheumatoid arthritis. *Arthritis Rheum.* 1990;**33**(4):591-601.
- 33 Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthritis Cartilage.* 2003;**11**(7):499-507.
- 34 Norman GR. Issues in the use of change scores in randomized trials. *J Clin Epidemiol.* 1989;**42**(11):1097-105.
- 35 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;**40**(2):171-8.
- 36 Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol.* 1992;**45**(12):1341-5.
- 37 Faucher M, Poiradeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. *Osteoarthritis Cartilage.* 2002;**10**(8):602-10.
- 38 Berrios-Rivera JP, Street RL, Jr., Garcia Popa-Lisseanu MG, Kallen MA, Richardson MN, Janssen NM, *et al.* Trust in physicians and elements of the medical interaction in patients with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum.* 2006;**55**(3):385-393.
- 39 Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis.* 1978;**37**(4):378-81.
- 40 Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain.* 1986;**27**(1):117-26.
- 41 Jaeschke R, Singer J, Guyatt GH. A comparison of seven-point and visual analogue scales. Data from a randomized trial. *Control Clin Trials.* 1990;**11**(1):43-51.
- 42 Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis.* 1987;**40**(12):1129-33.
- 43 van Laerhoven H, van der Zaag-Loonen HJ, Derkx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr.* 2004;**93**(6):830-5.
- 44 Grant S, Aitchison T, Henderson E, Christie J, Zare S, McMurray J, *et al.* A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest.* 1999;**116**(5):1208-17.
- 45 Brunier G, Graydon J. A comparison of two methods of measuring fatigue in patients on chronic haemodialysis: visual analogue vs Likert scale. *Int J Nurs Stud.* 1996;**33**(3):338-48.

Table 1. Baseline characteristics of patients involved in analyses.

	<i>Patients involved in analyses</i> <i>n=1218</i>
Age (years)	66.9 ±10.4
Sex (female)	854 (70.1%)
Weight (kg)	74.2 ± 13.9
Body mass index (kg/m²)	27.6 ± 4.7
Disease duration (years)	4.8 ± 5.7
Kellgren-Lawrence grade of osteoarthritis	
I	246 (20.2%)
II	530 (43.6%)
III	440 (36.2%)
Knee osteoarthritis	914 (75.0%)
Hip osteoarthritis	304 (25.0%)
NSAID intake during the past month	355 (29.2%)
Analgesic intake other than NSAID during the past month	699 (57.6%)
Symptomatic slow-action anti-arthrotic drug intake ⁽¹⁾	437 (36.0%)
Pain VAS (/10)	5.8 ± 1.7
Global assessment VAS (/10)	5.9 ± 1.9
WOMAC function subscale (/100)	43.7 ± 19.7

NSAID: non-steroidal anti-inflammatory drugs; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities osteoarthritis index.
(1) Chondroitin sulphate, diacerheine, or avocado/soybean unsaponifiables.
Results are presented as mean ±SD or n (%)

Table 2. Mean baseline score and mean change in score for all scales for the entire study population involved in analyses and the Likert 5, Likert 3 and VAS groups.

WOMAC function subscale		Individualized WOMAC multiplicative	Individualized WOMAC additive	WOMAC top 5
Study population (n=1218)	Mean baseline score	NR	NR	58.04 ±16.46 -15.95 ±16.70
	Mean change in score			
Likert 5 group (n=413)	Mean baseline score	27.64 ±13.08 -6.57 ±8.24	51.46 ±13.08 -4.37 ±5.64	55.80 ±16.01 -13.95 ±15.42
	Mean change in score			
Likert 3 group (n=406)	Mean baseline score	34.10 ±16.02 -9.67 ±11.48	55.18 ±13.87 -7.22 ±8.51	58.44 ±16.69 -17.15 ±17.14
	Mean change in score			
VAS group (n=399)	Mean baseline score	28.58 ±15.66 -8.49 ±9.99	51.70 ±17.79 -3.35 ±4.37	59.96 ±16.47 -16.80 ±17.36
	Mean change in score			

NR: not relevant; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities osteoarthritis index
Each score was linearly transformed to a 0 to 100 score.
Results are presented as mean ±SD.
Mean changes in scores were calculated between the baseline visit at week 0 (W0) and the final visit at week 4 (W4): W4-W0.

Table 3. Psychometric properties of individualized scales compared to those of the WOMAC function subscale.

WOMAC function subscale	Individualized WOMAC multiplicative		Individualized WOMAC additive		WOMAC top 5
	<i>Likert 5</i>	<i>Likert 3</i>	<i>Likert 5</i>	<i>Likert 3</i>	<i>VAS</i>
Construct validity					
<i>Spearman's correlation with:</i>					
WOMAC function subscale	1				
Pain (VAS)	0.93	0.95	0.93	0.93	0.79
Patient global assessment (VAS)	0.39	0.48	0.48	0.46	0.39
Practitioner global assessment (VAS)	0.41	0.46	0.45	0.44	0.34
	0.44	0.45	0.48	0.43	0.35
					0.81
					0.43
					0.46
					0.47
Internal consistency					
<i>Cronbach's alpha</i>	0.93				NE
[95% CI]	[0.92;0.93]				
Test-retest reliability					
<i>ICC</i>	0.91	0.92	0.93	0.93	0.94
[95% CI]	[0.89;0.92]	[0.90;0.93]	[0.92;0.94]	[0.91;0.94]	[0.93;0.95]
<i>mean change</i> * \pm <i>SD</i>	0.83		0.81		
	[0.74;0.90]	NA	[0.70;0.88]	NA	NA
	4.24 \pm 9.12		4.24 \pm 9.23		
Sensitivity to change					
<i>SRM</i>	0.80	0.84	0.85	0.86	0.77
[95% CI]	[0.68;0.92]	[0.74;0.97]	[0.76;0.96]	[0.74;0.97]	[0.65;0.90]
					[0.89;1.02]

ICC: intraclass correlation coefficient; NA: not available; NE: not estimable; SRM: standardized response mean; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities osteoarthritis index

For Spearman's correlations, all p-values were <0.0001

[95% CI]: 95% confidence intervals were estimated by the bootstrap method with 1000 replications

* **mean change in score between scores from baseline and scores of the reliability study.**

CAPTIONS OF ADDITIONAL FILES

Additional figure : Plots of severity against importance scores, for each item of the WOMAC function subscale for the VAS group (and Spearman's rho correlation coefficients (ρ) between the 2 scores).

Additional Table: Psychometric properties of the WOMAC short form and its individualized form compared to those of the WOMAC function subscale.

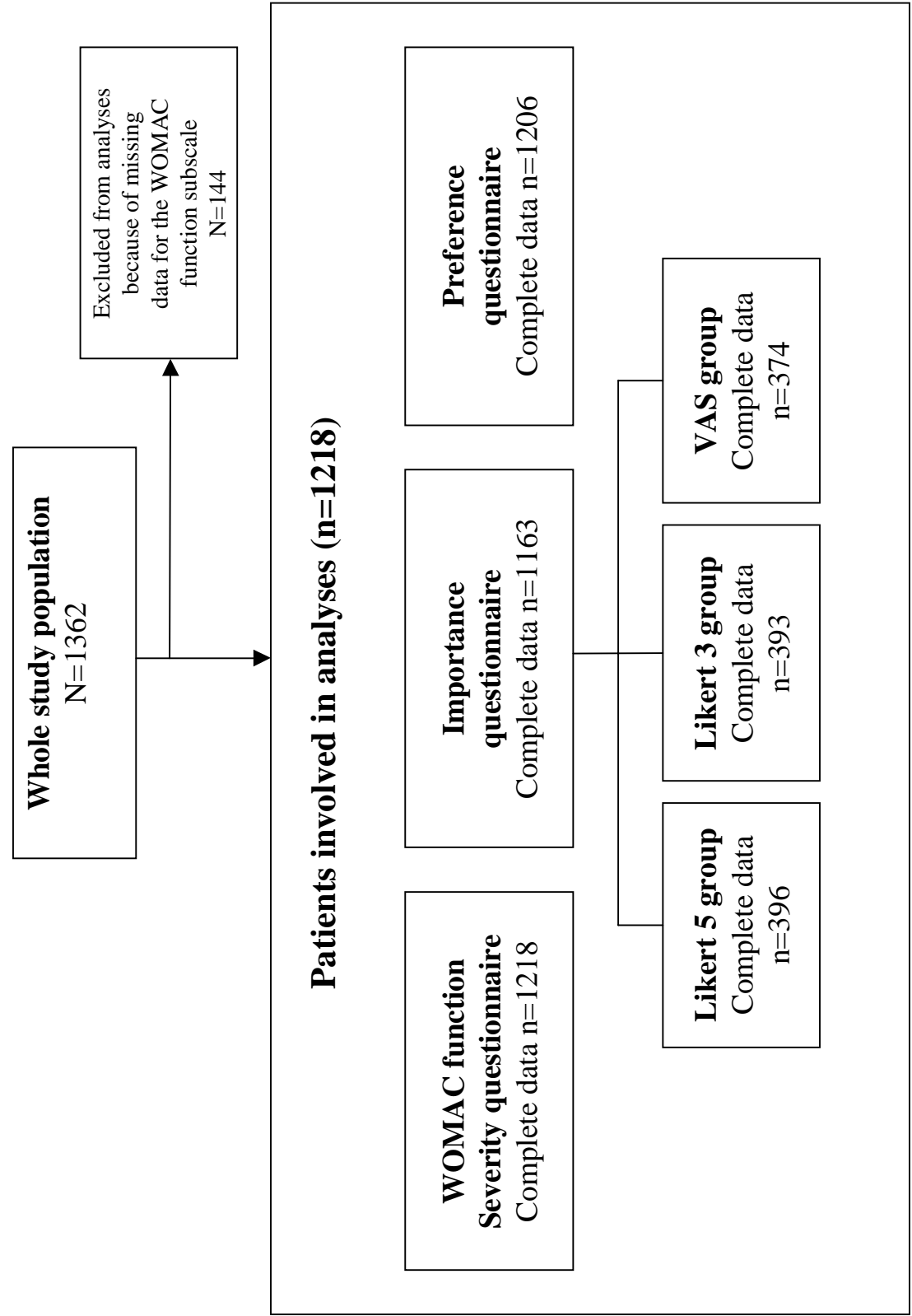
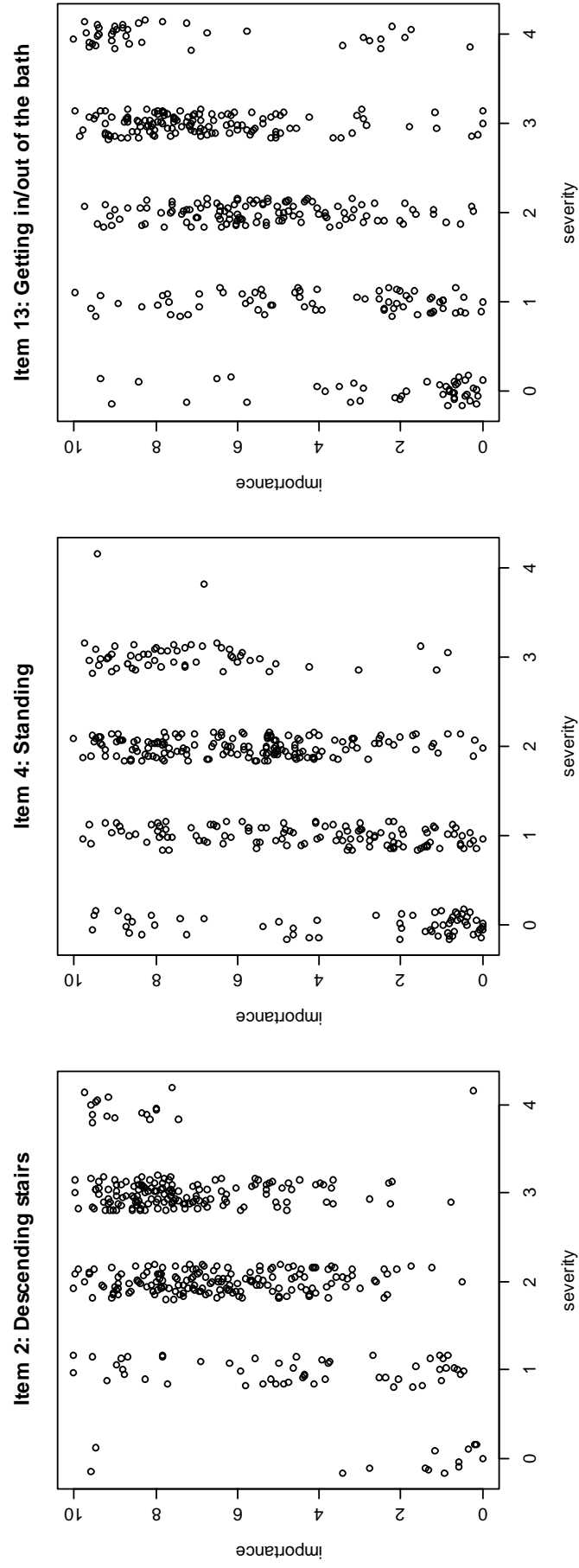
Fig. 1

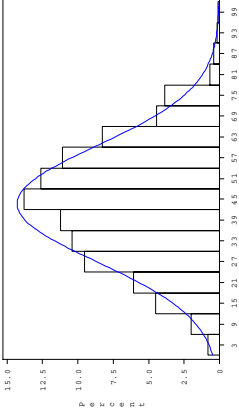
Fig. 2

Additional Table: Psychometric properties of the WOMAC short form and its individualised forms compared to those of the WOMAC function subscale.

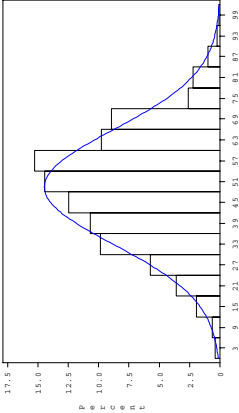
WOMAC function subscale		WOMAC short form	Individualized WOMAC short form multiplicative			Individualized WOMAC short form additive		
			Likert 5	Likert 3	VAS	Likert 5	Likert 3	VAS
Construct validity								
Spearman's correlation with:								
WOMAC function subscale		1	0.87	0.88	0.88	0.81	0.88	0.76
Pain (VAS)		0.46	0.42	0.48	0.48	0.38	0.47	0.40
Patient global assessment (VAS)		0.47	0.42	0.46	0.44	0.37	0.45	0.35
Practitioner global assessment (VAS)		0.49	0.46	0.44	0.46	0.41	0.42	0.35
Internal consistency								
Cronbach's alpha		0.93	0.82	0.83	0.86	0.85	0.83	0.87
[95% CI]		[0.92;0.93]	[0.79;0.85]	[0.79;0.85]	[0.84;0.88]	[0.82;0.87]	[0.79;0.85]	[0.85;0.89]
Test-retest reliability								
ICC		0.82	0.77	NA	NA	0.81	NA	NA
[95% CI]		[0.72;0.90]	[0.67;0.87]			[0.72;0.88]		
Sensitivity to change								
SRM		0.80	0.81	0.88	0.90	0.81	0.91	0.83
[95% CI]		[0.74;0.87]	[0.72;0.92]	[0.77;1.01]	[0.79;1.01]	[0.70;0.92]	[0.79;1.03]	[0.72;0.96]

ICC: intraclass correlation coefficient; NA: not available; SRM: standardized response mean; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities osteoarthritis index

Original WOMAC function subscale

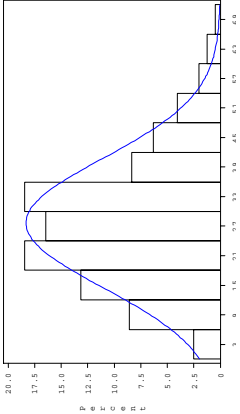


WOMAC short form

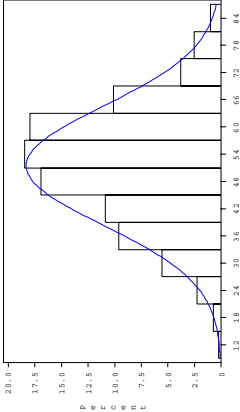


Individualized WOMAC function subscale

with multiplicative methods

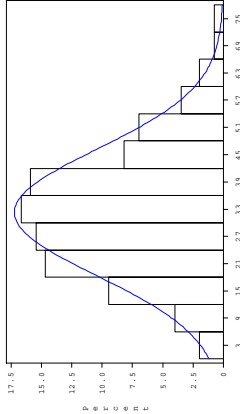


with additive methods

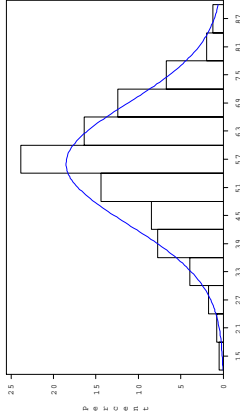


Individualized WOMAC short form

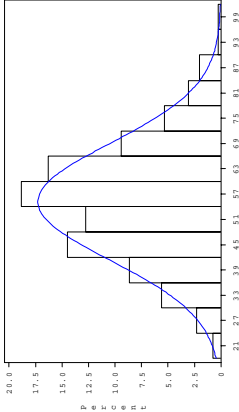
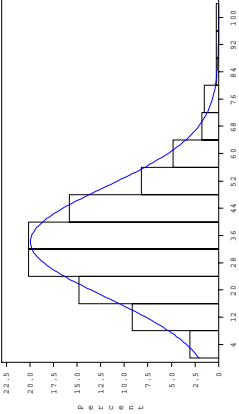
with multiplicative methods



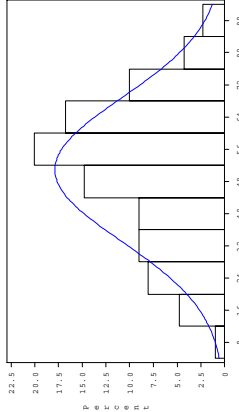
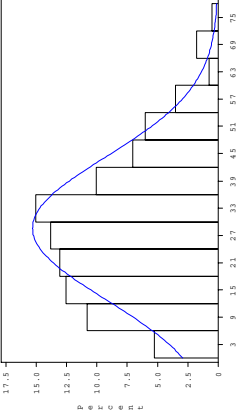
with additive methods



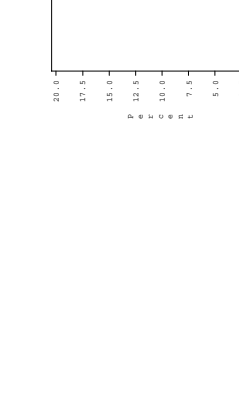
Likert 5 group



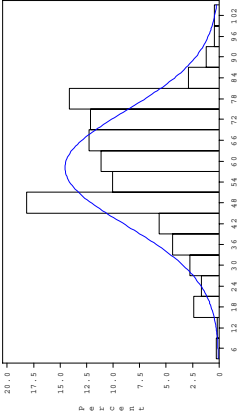
Likert 3 group



VAS group



WOMAC Top 5



Additional figure 2

